BMJ Open

Validation of intellectual disability through hospital morbidity records using a population-based database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019113
Article Type:	Research
Date Submitted by the Author:	12-Aug-2017
Complete List of Authors:	Bourke, Jenny; Telethon Kids Institute, Epidemiology Wong, Kingsley; Telethon Kids Institute Leonard, Helen
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Research methods
Keywords:	intellectual disability, hospital morbidity coding, linked data

Validation of intellectual disability through hospital morbidity records using a population-based database

Jenny Bourke, MPH[1] Kingsley Wong, MBBS[1] Helen Leonard, MBChB[1]

[1] Telethon Kids Institute, University of Western Australia, Perth, Western Australia, Australia

Corresponding author

Helen Leonard

helen.leonard@telethonkids.org.au

MOB:+61 419956946

Telethon Kids Institute

100 Roberts Rd, Subiaco

WA 6008 Australia

Word count=2580

Abstract

Objectives: To investigate how well intellectual disability can be ascertained using hospital morbidity data compared with a population-based data source.

Design, Setting and Participants: All children born 1983-2010 with a hospital admission in the Western Australian Hospital Morbidity Data System (HMDS) were linked with the Western Australian Intellectual Disability (IDEA) database. The ICD hospital codes consistent with intellectual disability were also identified.

Main Outcome measures: The characteristics of those children identified with intellectual disability through either or both sources were investigated.

Results: Of the 488,905 individuals in the study, 10,218 (2.1%) were identified with intellectual disability in either IDEA or HMDS with 1,435 (14.0%) individuals identified in both databases, 8,305 (81.3%) unique to the IDEA database and 478 (4.7%) unique to the HMDS dataset only. Of those unique to the HMDS dataset, about a quarter (n=124) had died before one year of age and most of these (75%) before one month. Children with intellectual disability who were also coded as such in the HMDS data were more likely to be aged under one year, female, non-Aboriginal and have a severe level of intellectual disability, compared with those not coded in the HMDS data. The sensitivity of using HMDS to identify intellectual disability was 14.7%, whilst the specificity was much higher at 99.9%.

Conclusion: Hospital morbidity data are not a reliable source for identifying intellectual disability within a population and epidemiological researchers need to take these findings into account in their study design.

Strengths and limitations of this study

- The greatest strength of this study was the availability of a population-based source of intellectual disability (ID).
- The state-wide data linkage system allowed this database to be linked to other population datasets such as hospital morbidity.
- The study was able to identify characteristics of children known to have ID by whether or not they were not identified with ID within hospital morbidity data.
- One limitation is that for some conditions associated with ID not all children will necessarily meet criteria for intellectual disability.
- The ICD9/10 coding system is limited in its provision of delineation of some genetic syndromes, however the integration of Orphanet coding into ICD-11 will allow many more genetic ID syndromes to be specifically identified in hospital morbidity data.

Introduction

Intellectual disability (ID) is characterised by globally impaired cognitive functioning and significant deficits in adaptive functioning, manifest before the age of 18 years.[1] Comorbid medical or psychiatric conditions are common in people with ID,[2] leading to increased hospitalisations. The increased risk of admission has been shown to range from two-fold for those with ID associated with autism up to ten-fold for those with severe ID.[3] For conditions typically managed through ambulatory (out-patient) care, people with ID have been shown to have a six-fold increase in risk of hospitalisations compared to those without ID.[4] Epilepsy is one of the most common health conditions in this population with a prevalence of around 20%[2, 5] and is one of the main reasons for hospital admission.[4] Specific disorders consistent with ID such as Down syndrome are often associated with multiple medical conditions (e.g. cardiac defects, ear disease, respiratory infections) which often require hospitalisation.[6] Mental health disorders are also more prevalent in individuals with ID [7] and hospitalisation is common.[4, 8]

Children and young adults with ID however, form a heterogeneous group and reliable population-based information on the prevalence of intellectual disability is often difficult to obtain. On account of the high likelihood of hospitalisations[3] health data have often been used as a source of ascertainment for prevalence studies of ID, as have other administrative datasets relating to education or service provision as well as household sampling.[9] Using such sources the prevalence varies widely and may be unreliable, also because of the use of different diagnostic systems to identify

the presence of ID. Overall, higher estimates were found in studies using household sampling compared with hospital or administrative data, and in those studies where ID was diagnosed using psychological assessment, compared with ICD or DSM codes.[9] Studies have also used health related datasets including insurance claims to identify ID and investigate specific causes of hospitalisation in this population.[4, 10]

In Western Australia the IDEA database is a population-based register of children with ID, which uses ascertainment from both disability service providers and education sources.[11] It is a research infrastructure that can be linked to other population datasets such as hospital morbidity data.[12] The current study aims to investigate how well the Western Australian Hospital Morbidity Data System (HMDS), which contains all admissions to private and public hospitals, recorded ID compared to the IDEA database and thus assess the usefulness of hospitalisation data as a source of ID status.

Methods

The study cohort was restricted to children and young adults born between 1983 and 2010 and who were identified with ID in either the HMDS or the IDEA database over this period. Individuals were defined as having an intellectual disability in the HMDS if they were assigned any of the following International Classification of Diseases (ICD) diagnostic codes during hospitalisation: Mental retardation (ICD-9-CM 317-319; ICD-10-AM F70-F79), Down syndrome [Trisomy 21] (ICD-9-CM 758.0; ICD-10-AM Q90.0-Q90.2, Q90.9), Edwards/Patau syndrome [Trisomy 18/13] (ICD-9-CM 758.1, 758.2; ICD-10-AM Q91.0-Q91.7), Trisomy 9/8 (ICD-9-CM 758.5; ICD-10-AM

Q92.0-Q92.5), Chromosomal deletions (ICD-9-CM 758.3; ICD-10-AM Q93.3-Q93.5), Fragile X syndrome (ICD-9-CM 759.83; ICD-10-AM Q99.2), Neurofibromatosis (ICD-9-CM 237.7; ICD-10-AM Q85.0), Tuberous sclerosis (ICD-9-CM 759.5; ICD-10-AM Q85.1), Prader-Willi syndrome (ICD-9-CM 759.81; ICD-10-AM Q87.14), and Marfan syndrome (ICD-9-CM 759.82; ICD-10-AM Q87.4). Individuals diagnosed with an intellectual disability in the IDEA database, considered the "gold standard" for ID diagnosis in the Western Australian population, were linked to the HMDS data. Maternal race (Aboriginal, non-Aboriginal), gender (male or female) and date of birth were obtained by linkage to the Midwives' Notification System. Information on deaths was obtained by linkage to the WA Mortality database and children and those who had died before one year of age were identified.

Age at admission (<1, 1-2, 3-5, 6-12 and >12 years), gender (male, female), race (non-Aboriginal, Aboriginal) and level of ID (mild or moderate, severe) of individuals with an intellectual disability in the IDEA dataset were compared between those who were and were not identified in the HMDS. The main cause of intellectual disability as recorded in the IDEA database using the Heber codes,[13] was further grouped into broad categories based on biomedical or other causes[14] in order to investigate whether the cause of ID differed between those identified and not identified with ID in the HMDS dataset. Categorical variables were reported as proportions and compared using the Pearson's chi-squared test for independence. Analyses were performed using Stata 13.1 (StataCorp, College Station, Texas, USA).

Results

A total of 1,548,478 records representing admissions for 488,905 individuals were identified. Among them, 10,218 (2.1%) were identified as having an ID and 478,687 (97.9%) cases as not having ID in either the HMDS or the IDEA database. Of those who were diagnosed with ID, 1,435 (14.0%) were identified in both, 8,305 (81.3%) were unique to the IDEA database and 478 (4.7%) were unique to the HMDS dataset (Figure 1). Death before the age of one year had occurred in 160 / 10,218 (1.5%) of the individuals identified with ID in either source with the majority (n=124, 78%) of these unique to HMDS. Limited to those who survived past one year of age, the sensitivity of using HMDS to identify ID was 14.6%, whilst the specificity was much higher at 99.9%. The positive and negative predictive values were 79.9% and 98.3% respectively.

We compared the characteristics of the 9,704 individuals who were registered in the IDEA database and thus known to have an ID, survived past one year of age and were admitted to hospital by whether they were identified with ID in HMDS (Table 1). Children with ID who were also coded with ID in the HMDS data were more likely to be less than one year of age at first admission compared with children with ID not coded in the HMDS data (79.2% vs 68.0%). They were also more likely to be female (44.6% vs 33.8%), be non-Aboriginal (92.2% vs 85.7%) and have a severe level of ID (21.6% vs 6.2%).

Children in the IDEA database with a biomedical cause of their ID were more likely to have also been coded with ID in the HMDS dataset (Table 2). The causes most likely to have been identified with ID were Down syndrome (94.2%), Tuberous sclerosis (90.6%), Prader-Willi syndrome (87.0%), Neurofibromatosis (70.6%), muscular dystrophy (57.1%) and Fragile X (51.6%). Those least likely to have been identified with ID were those with an unassessed cause (2.7%), autism (3.0%) Asperger's (3.9%), foetal alcohol syndrome (8.0%) and other associated conditions such as intrauterine growth restriction (2.9%) and prematurity (5.6%) (Table 2). Additionally, 30% of children who had been identified with any epilepsy diagnosis, regardless of their main cause of ID diagnosis, were found to be identified with ID in the hospital dataset (not shown in Table).

Children identified with ID in the HMDS dataset who were not in the IDEA database and had survived one year were investigated according to the codes used to identify ID in HMDS (Table 3). The majority of those not in IDEA had been assigned an ICD code aligned to mental retardation (n=138, 39.0%), Neurofibromatosis (n=79, 22.3%) or Down syndrome (n=45, 12.7%) (Table 3). Among the 124 (25.9%) individuals who had died before one year of age, 75% had died before one month, and the majority of diagnoses included Trisomy 18/13 (n=80, 64.5%), Down syndrome (n=25, 20.2%) or Trisomy 8/9 (n=10, 8.1%).

Discussion

Data from Western Australia suggest that hospital morbidity data may be an inadequate source of identification of intellectual disability in epidemiological

studies with a sensitivity of only 14%. After removing children who died before one year of age, intellectual disability of syndromic or monogenic aetiology such as that associated with Down syndrome, neurofibromatosis and Fragile X syndrome was most likely also to be identified in hospital sources and ID of unknown cause least likely to be identified. Females and children under one year were also more likely to be identified while Aboriginal children and those with a mild-moderate level of intellectual disability were less likely to be identified.

The greatest strength of this study was the availability of a population source of ID, the IDEA database which has used both disability service use and education sources to maintain high ascertainment over the last thirty years.[15] It has already been used as a data source for multiple data linkage studies investigating determinants[16-18] and outcomes[3, 19] associated with intellectual disability. One limitation is that there are several conditions, where only a percentage of children have an intellectual disability in contrast to conditions like Down syndrome where almost all children are affected. However for the purposes of this study we still elected to use the codes for these diagnoses to identify ID in the HMDS. Thus by doing this and assigning ID status to all children with these conditions in hospital morbidity records we could have overestimated the number with ID. For example, intellectual disability is diagnosed in approximately half of individuals with tuberous sclerosis[20] and whilst almost all of those with Prader Willi syndrome will have cognitive deficits, up to 40% may fall within the borderline range.[21] About a third of children with neurofibromatosis have been reported to have general learning difficulties associated with borderline or lower IQ[22] and children with Marfan syndrome may only have a slightly increased

risk of intellectual disability.[23] Children diagnosed with autism spectrum disorder have been found to have an ID in approximately 30%- 60% of cases although this proportion has been shown to be decreasing in more recent years.[17, 24, 25] The effect of removing these conditions from our HMDS search list would have been to slightly increase the sensitivity and positive predictive value of using HMDS to identify ID.

Children with a cause of ID commonly known to be associated with ID, such as Down syndrome or Prader Willi, were most likely to be identified in the hospital data, unlike those for whom no clear cause had been recorded in the IDEA Database. The inability of ICD codes to specifically identify relatively rare conditions is also problematic if relying on such identification of ID. For example, Williams syndrome, known to be highly associated with ID,[26] is identified with a Q89.8 ICD-10 code which is in itself not specific for Williams syndrome and was not used in our search strategy as it would also identify children possibly without ID such as those with Stickler syndrome. Perhaps as a consequence, children with Williams syndrome were poorly identified as ID in the hospital codes, with only 16% of children being coded as such. Recent versions of ICD-10-AM provide a finer delineation of genetic syndromes and thus allow better differentiation of syndromes with ID from those without the condition. The integration of Orphanet coding into ICD-11 will allow many more genetic ID syndromes to be specifically identified in hospital morbidity data.[27] This has become a matter of urgency given the accelerated identification of these genetic causes over the last decade and particularly since the introduction of next generation sequencing.[28-30]

Many children who would be expected to develop ID by virtue of their diagnosis experience serious and life-threatening comorbidities and as a consequence may die early. As we have shown, about a third of those not identified in the IDEA database had died, nearly three quarters before one month of age and the majority by one year. In these cases it would be unlikely that families would have sought registration for disability services before their child died and hence they would not have been included within the IDEA database.

We found one Canadian study which had used hospital morbidity codes to identify ID in at least one patient record in order to form their cohort, but had found that as many as half of the multiple records for these individuals did not code ID as a comorbidity in the hospital morbidity system.[8] It was therefore likely that other individuals with ID had been missed from their cohort due to inconsistent coding of ID as a comorbidity. The authors acknowledged that, similar to our own findings, it was likely that those who had been identified with ID were more severe. In a later study they estimated population prevalence of ID by identifying individuals with ID using the same ICD codes related to ID in a number of different health administrative datasets.[31] Using their broadest capture algorithm they found an overall prevalence of 8/1000 and a prevalence of 14.2/1000 in young adults aged 18-24 years,[31] not too dissimilar from our own estimate of 17/1000 in a similarly aged population.[15] These prevalence estimates based on health datasets certainly provided better ascertainment than the 14% capture using our own hospital morbidity codes but the ascertainment is still likely less complete than our population ascertainment using the IDEA database. Linked data studies in New South Wales, Australia have provided

further evidence of the need for multiple sources of ascertainment of ID.[32] Using ICD codes for ID within health datasets, as well as disability services, birth and mortality linkages, the authors found an overall prevalence of 0.6% (or 6/1000) considerably less than that found in our IDEA database.

Practical considerations for clinical care would suggest that hospital coding which does not include reference to intellectual disability as a comorbidity may impact on the way in which service is delivered to this particularly vulnerable population. Better coding practices for ID would enable researchers to investigate directly whether care or procedures are compromised for individuals with ID and facilitate the development of ID-related policies and service planning. The hospital experiences for people with ID, who we know experience higher rates of hospitalisation than the rest of the population,[3] have been described as relying heavily on carers for in-hospital patient assistance with failure to provide appropriate care, and lack of knowledge and discharge planning by medical staff.[33] The reliance on hospital morbidity data, as well as other administrative datasets, to identify ID in a population for research purposes, particularly prevalence estimates, has been shown to provide varied results.[9] Overall, we would not recommend that researchers use hospital morbidity datasets alone as a source of identification of intellectual disability.

Conclusion

Through linkage to a hospital morbidity dataset, this study has shown that hospital data does not adequately identify individuals with ID when compared with the

population-based IDEA database. A high proportion of those uniquely identified in hospital morbidity data had died early or alternatively they had a condition not necessarily associated with ID. It is important for hospital codes to reflect the ID status of patients, primarily for the benefit of recognizing their specific needs, but also for improvement of ascertainment of ID through this source. Clearly with such a high proportion of individuals not being recognized with ID, coding practices which identify ID need to be better implemented.

Acknowledgements

The authors acknowledge the Disability Services Commission, the Telethon Kids Institute, the Western Australia Department of Education, the Catholic Education Office and the Association of Independent Schools of Western Australia for assistance with data collection for the IDEA database.

Funding Statement

This study was partly funded by an Australian National Health and Medical Research Council (NHMRC) Program Grant #572742.

Competing Interests

We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Author Contributions

All authors contributed to the initial design of the manuscript. JB and HL were responsible for the drafting of the paper. KW was responsible for analysis and contributed to the writing of the final draft. All authors contributed to the final writing of the paper and checked for important intellectual content.

References

1. Schalock R, Borthwick-Duffy S, Bradley V, Buntinx W, Coulter D, Craig E, et al. Intellectual disability: Definition, classification, and systems of supports. 11th ed. Washington, DC : American Association on Intellectual and Developmental Disabilities; 2010.

2. Oeseburg B, Dijkstra GJ, Groothoff JW, Reijneveld SA, Jansen DE. Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review. Intellect Dev Disabil. 2011;49:59-85.

3. Bebbington A, Glasson E, Bourke J, de Klerk N, Leonard H. Hospitalisation rates for children with intellectual disability or autism born in Western Australia 1983-1999: a population-based cohort study. BMJ open. 2013;3.

4. Balogh R, Brownell M, Ouellette-Kuntz H, Colantonio A. Hospitalisation rates for ambulatory care sensitive conditions for persons with and without an intellectual disability--a population perspective. J Intellect Disabil Res. 2010;54:820-32.

5. Bowley C, Kerr M. Epilepsy and intellectual disability. J Intellect Disabil Res. 2000;44 (Pt 5):529-43.

6. Fitzgerald P, Leonard H, Pikora TJ, Bourke J, Hammond G. Hospital admissions in children with Down syndrome: experience of a population-based cohort followed from birth. PloS one. 2013;8:e70401.

7. Emerson E, Hatton C. Mental health of children and adolescents with intellectual disabilities in Britain. Br J Psychiatry. 2007;191:493-9.

8. Balogh RS, Hunter D, Ouellette-Kuntz H. Hospital Utilization among Persons with an Intellectual Disability, Ontario, Canada, 1995–2001. J Appl Res Intellect. 2005;18:181-90.

9. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. Res Dev Disabil. 2011;32:419-36.

10. Büchele G, Becker C, Cameron I, Auer R, Rothenbacher D, König H, et al. Fracture risk in people with developmental disabilities: results of a large claims data analysis. Osteoporosis Int. 2017;28:369-75.

11. Petterson B, Leonard H, Bourke J, Sanders R, Chalmers R, Jacoby P, et al. IDEA (Intellectual Disability Exploring Answers): a population-based database for intellectual disability in Western Australia. Ann Hum Biol. 2005;32:237-43.

12. Holman CDAJ, Bass AJ, Rouse IL, Hobbs MST. Population-based linkage of health records in Western Australia: development of a health services research linked database. Aust NZ J Publ Heal. 1999;23:453-9.

13. Heber R. A manual on terminology and classification in mental retardation. Am J Ment Def. 1961;64.

14. Yeargin‐Allsopp M, Murphy CC, Cordrro JF, Decoufle P, Hollowell JG. Reported biomedical causes and associated medical conditions for mental retardation among 10‐year‐old children, metropolitan Atlanta, 1985 to 1987. Dev Med Child Neurol. 1997;39:142-9.

15. Bourke J, de Klerk N, Smith T, Leonard H. Population-Based Prevalence of Intellectual Disability and Autism Spectrum Disorders in Western Australia: A Comparison With Previous Estimates. Medicine. 2016;95:e3737.

16. Leonard H, Glasson E, Bebbington A, Hammond G, Croft D, Pikora T, et al. Application of Population-Based Linked Data to the Study of Intellectual Disability and Autism. In: Urbano RC, editor. International Review of Research in Developmental Disabilities. MA, USA: Academic Press; 2013. p. 281-327.

17. Leonard H, Glasson E, Nassar N, Whitehouse A, Bebbington A, Bourke J, et al. Autism and intellectual disability are differentially related to sociodemographic background at birth. PloS one. 2011;6:e17875.

18. Leonard H, Nassar N, Bourke J, Blair E, Mulroy S, deKlerk N, et al. Intrauterine growth and subsequent intellectual disability in a ten-year population cohort of children in Western Australia. Am J Epidemiol. 2007;165:S19-S.

19. Bourke J, Nembhard WN, Wong K, Leonard H. Twenty-Five Year Survival of Children with Intellectual Disability in Western Australia. J Pediatr. 2017; in press

20. Curatolo P, Bombardieri R, Jozwiak S. Tuberous sclerosis. The Lancet. 2008; 372:657-68.

21. Cassidy SB, Driscoll DJ. Prader–Willi syndrome. Eur J Hum Genet. 2009;17:3-13.

22. Hyman SL, Arthur E, North KN. Learning disabilities in children with neurofibromatosis type 1: subtypes, cognitive profile, and attention-deficit-hyperactivity disorder. Dev Med Child Neurol. 2006;48:973-7.

23. Dietz H. Marfan Syndrome Seattle (WA): University of Washington, Seattle; 2001 [updated 2 Feb, 201711 July, 2017]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1335/.

24. Baio J. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. MMWR Surveill Summ. 2014;63:1-21.

25. Nassar N, Dixon G, Bourke J, Bower C, Glasson E, de Klerk N, et al. Autism spectrum disorders in young children: effect of changes in diagnostic practices. Int J Epidemiol. 2009;38:1245-54.

26. Greer MK, Brown FR, Pai GS, Choudry SH, Klein AJ. Cognitive, adaptive, and behavioral characteristics of Williams syndrome. Am J Med Genet. 1997;74:521-5.

27. Walker CE, Mahede T, Davis G, Miller LJ, Girschik J, Brameld K, et al. The collective impact of rare diseases in Western Australia: an estimate using a population-based cohort. Genet Med. 2017;19:546-52.

28. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511:344-+.

29. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2016;17:9-18.

30. Valdez R, Grosse SD, Khoury MJ. The need for a next-generation public health response to rare diseases. Genet Med. 2017;19:489-490.

31. Lin E, Balogh R, Cobigo V, Ouellette-Kuntz H, Wilton AS, Lunsky Y. Using administrative health data to identify individuals with intellectual and developmental disabilities: a comparison of algorithms. J Intellect Disabil Res. 2013;57:462-77.

32. Reppermund S, Srasuebkul P, Heintze T, Reeve R, Dean K, Emerson E, et al. Cohort profile: a data linkage cohort to examine health service profiles of people with intellectual disability in New South Wales, Australia. BMJ open. 2017;7:e015627.

33. Iacono T, Bigby C, Unsworth C, Douglas J, Fitzpatrick P. A systematic review of hospital experiences of people with intellectual disability. BMC Health Serv Res. 2014;14:505.

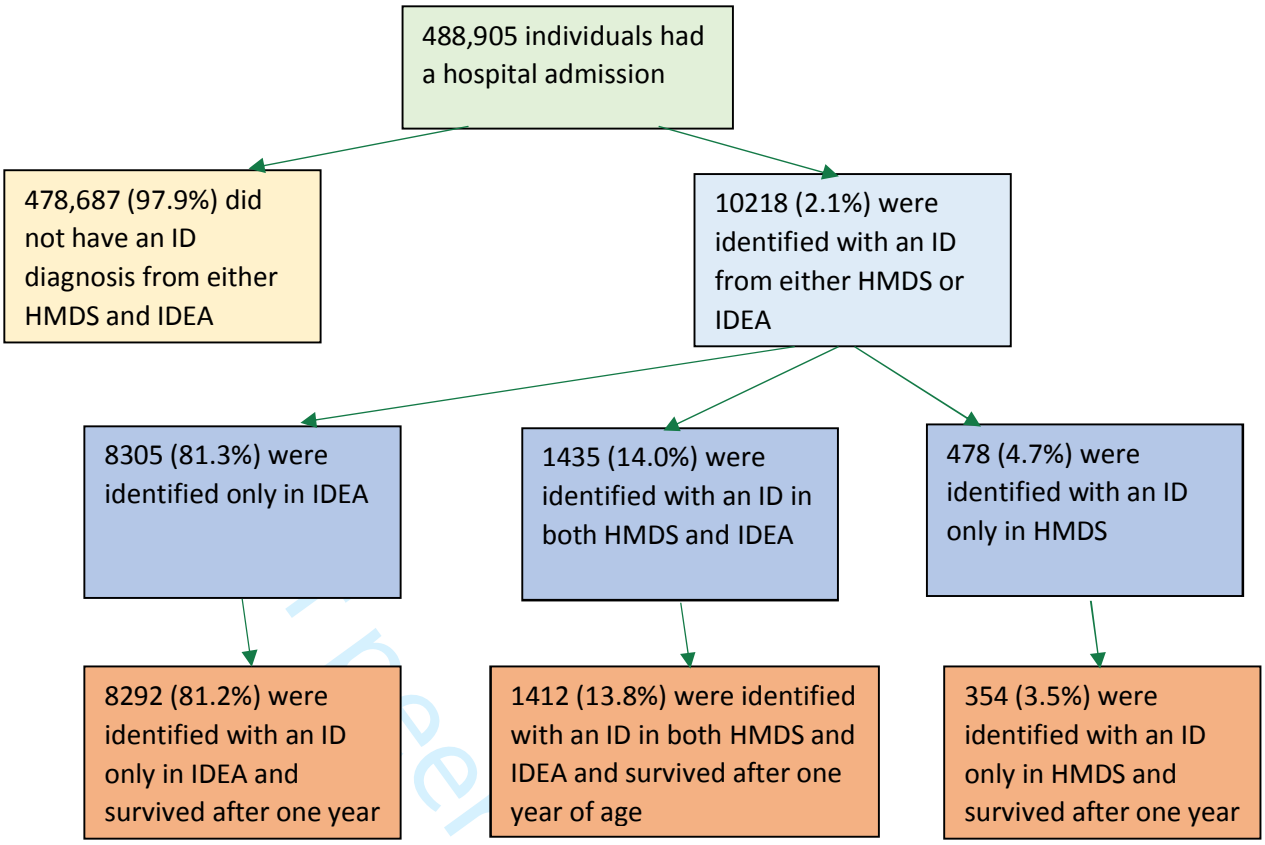

Figure 1: Identification of intellectual disability (ID) in children born 1983-2010 and hospitalised in Western Australia using linkage to the IDEA database and the hospital morbidity data system (HMDS)

Table 1: Characteristics of children born between 1983 and 2010 in Western Australia and survived past one year of age, who were identified with intellectual disability (ID) through the IDEA database and admitted to hospital, according to their ID diagnosis status in the Hospital Morbidity Data System (HMDS) database

	ID diagnosis status in HMDS, N (%)			
Characteristic	**Yes**	**No**	**Total**	**P-value***
Age at first admission (years)				
<1	1,119 (79.2)	5,636 (68.0)	6,755 (69.6)	
1-2	177 (12.3)	1,256 (15.1)	1,433 (14.7)	
3-5	54 (3.8)	714 (8.6)	768 (7.9)	<0.01
6-12	36 (2.6)	436 (5.3)	472 (4.9)	
>12	26 (1.8)	250 (3.0)	276 (2.8)	
Gender				
Male	782 (55.4)	5,489 (66.2)	6,271 (64.6)	<0.01
Female	630 (44.6)	2,803 (33.8)	3,433 (35.4)	
Race				
Non-Aboriginal	1,302 (92.2)	7,106 (85.7)	8,408 (86.6)	<0.01
Aboriginal	110 (7.8)	1,186 (14.3)	1,296 (13.4)	
Level of ID				
Mild or moderate ID	1,107 (78.4)	7,776 (93.8)	8,883 (91.5)	<0.01
Severe ID	305 (21.6)	516 (6.2)	821 (8.5)	
Total	1,412 (100)	8,292 (100)	9,704 (100)	

* Pearson's chi-squared test for independence

HMDS, Hospital Morbidity Data System; ID, intellectual disability

Table 2: Children in the IDEA database who survived to one year of age and identified/ not identified with ID within the HMDS dataset, by the cause of intellectual disability

Cause of ID	In IDEA and identified with ID in HMDS		In IDEA and not identified with ID in HMDS		Total
1. PRENATAL CONDITIONS	n	%	n	%	n
Genetic or Chromosomal:					
Down Syndrome	589	94.2	36	5.8	625
Tuberous Sclerosis	29	90.6	3	9.4	32
Prader Willi Syndrome	20	87.0	3	13.0	23
Neurofibromatosis	12	70.6	5	29.4	17
Muscular Dystrophy	4	57.1	3	42.9	7
Fragile X	16	51.6	15	48.4	31
Other Chromosomal	59	45.0	72	55.0	131
X-linked inheritance	4	36.4	7	63.6	11
Metabolic (possible)	9	29.0	22	71.0	31
Myotonic Dystrophy	3	27.3	8	72.7	11
Syndrome Grouped	45	26.5	125	73.5	170
Mucopolysaccharidosis	1	25.0	3	75.0	4
Autosomal	21	23.9	67	76.1	88
Prenatal aetiology	8	18.2	36	81.8	44
Williams syndrome	5	16.1	26	83.9	31
Neurodegenerative disorders	1	11.1	8	88.9	9
Sex Chromosome	2	9.5	19	90.5	21
Mitochondria	1	7.7	12	92.3	13
Metabolic	1	5.9	16	94.1	17
Teratogenic:					
Cytomegalic Inclusion congenital	12	50.0	12	50.0	24
Other potential teratogens	4	16.7	20	83.3	24
Other prenatal infections	1	9.1	10	90.9	11
Potential Foetal alcohol syndrome	7	8.0	81	92.1	88
CNS and Other Birth Defects:					
Unspecified Neurological	32	42.7	43	57.3	75
Congenital hypothyroidism	1	25.0	3	75.0	4
Spina Bifida Meningocoele	3	25.0	9	75.0	12

Unknown Prenatal	51	22.6	175	77.4	226
Microcephaly	7	17.5	33	82.5	40
CNS: Malformations of Gyri	4	17.4	19	82.6	23
Hydrocephalus	4	16.7	20	83.3	24
Macrocephaly	3	16.7	15	83.3	18
Cranial anomalies	6	16.2	31	83.8	37
CNS Malformations	6	10.2	53	89.8	59
2. PERINATAL CONDITIONS					
Hypoxic Ischaemic Encephalopathy	27	29.0	66	71.0	93
Perinatal: Neonatal	2	28.6	5	71.4	7
3. POSTNEONATAL CONDITIONS					
Post Natal Asphyxia	13	44.8	16	55.2	29
Postnatal Injury	23	31.5	50	68.5	73
Postneonatal infection	21	29.6	50	70.4	71
Intracranial Neoplasm	2	28.6	5	71.4	7
4. NO DEFINED CAUSE					
Associated with Epilepsy	44	24.2	138	75.8	182
Cultural Familial IH	29	20.4	113	79.6	142
Associated with Coexisting disability	2	20.0	8	80.0	10
Associated with Psychotic Disorder	4	14.3	24	85.7	28
Associated Maternal medical condition	4	10.0	36	90.0	40
No defined cause *(Functional reaction alone)*	66	8.7	689	91.3	755
Other Developmental Disorders	3	8.3	33	91.7	36
Familial Unspecified	20	6.3	300	93.8	320
Associated with Psychosocial factors	2	6.3	30	93.8	32
Prematurity	9	6.3	133	93.7	142
Multiple Birth	2	5.0	38	95.0	40
Aspergers	1	3.9	25	96.2	26
Autism	42	3.0	1,342	97.0	1,384
Intrauterine growth restriction	1	2.9	34	97.1	35
Unassessed	114	2.7	4,103	97.3	4,217
Total	1,412	14.6	8,292	85.4	9,704

Table 3: Children born between 1983 and 2010 in Western Australia and were identified with intellectual disability (ID) in the Hospital Morbidity Data System (HMDS) database but not identified in the IDEA database, by death status and ID diagnosis in HMDS

ID diagnosis in HMDS	Died under one year	Alive after one year	Total
	n (%)	n (%)	n (%)
Mental retardation	3 (2.4)	138 (39.0)	141 (29.5)
Down syndrome	25 (20.2)	45 (12.7)	70 (14.6)
Trisomy 18/13	80 (64.5)	5 (1.4)	85 (17.8)
Trisomy 8/9	10 (8.1)	12 (3.4)	22 (4.6)
Chromosomal deletion	5 (4.0)	16 (4.5)	21 (4.4)
Fragile X	0	1 (0.3)	1 (0.2)
Neurofibromatosis	0	79 (22.3)	79 (16.5)
Tuberous sclerosis	0	17 (4.8)	17 (3.6)
Prader-Willi syndrome	0	6 (1.7)	6 (1.3)
Marfan syndrome	1 (0.8)	35 (9.9)	36 (7.5)
Total	124	354	478

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 5
Methods			
Study design	4	Present key elements of study design early in the paper	Pages 5,6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Pages 5,6
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Page 6
Bias	9	Describe any efforts to address potential sources of bias	n/a
Study size	10	Explain how the study size was arrived at	Page 5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Page 6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 6
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6
		(*c*) Explain how missing data were addressed	n/a
		(*d*) If applicable, explain how loss to follow-up was addressed	n/a
		(*e*) Describe any sensitivity analyses	n/a
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Page 7
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) Summarise follow-up time (eg, average and total amount)	n/a
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Page 7
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n/a
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 8,9
Limitations			Page 9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Pages 10,11
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Validation of intellectual disability coding through hospital morbidity records using an intellectual disability population-based database in Western Australia

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019113.R1
Article Type:	Research
Date Submitted by the Author:	09-Oct-2017
Complete List of Authors:	Bourke, Jenny; Telethon Kids Institute, Epidemiology Wong, Kingsley; Telethon Kids Institute Leonard, Helen
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Research methods
Keywords:	intellectual disability, hospital morbidity coding, linked data

Validation of intellectual disability coding through hospital morbidity records using an intellectual disability population-based database in Western Australia

Jenny Bourke, MPH[1] Kingsley Wong, MBBS[1] Helen Leonard, MBChB[1]

[1] Telethon Kids Institute, University of Western Australia, Perth, Western Australia, Australia

Corresponding author

Helen Leonard

helen.leonard@telethonkids.org.au

MOB:+61 419956946

Telethon Kids Institute

100 Roberts Rd, Subiaco

WA 6008 Australia

Word count=2580

Abstract

Objectives: To investigate how well intellectual disability can be ascertained using hospital morbidity data compared with a population-based data source.

Design, Setting and Participants: All children born 1983-2010 with a hospital admission in the Western Australian Hospital Morbidity Data System (HMDS) were linked with the Western Australian Intellectual Disability (IDEA) database. The ICD hospital codes consistent with intellectual disability were also identified.

Main Outcome measures: The characteristics of those children identified with intellectual disability through either or both sources were investigated.

Results: Of the 488,905 individuals in the study, 10,218 (2.1%) were identified with intellectual disability in either IDEA or HMDS with 1,435 (14.0%) individuals identified in both databases, 8,305 (81.3%) unique to the IDEA database and 478 (4.7%) unique to the HMDS dataset only. Of those unique to the HMDS dataset, about a quarter (n=124) had died before one year of age and most of these (75%) before one month. Children with intellectual disability who were also coded as such in the HMDS data were more likely to be aged under one year, female, non-Aboriginal and have a severe level of intellectual disability, compared with those not coded in the HMDS data. The sensitivity of using HMDS to identify intellectual disability was 14.7%, whilst the specificity was much higher at 99.9%.

Conclusion: Hospital morbidity data are not a reliable source for identifying intellectual disability within a population and epidemiological researchers need to take these findings into account in their study design.

Strengths and limitations of this study

- The greatest strength of this study was the availability of a population-based source of intellectual disability (ID).
- The state-wide data linkage system allowed this database to be linked to other population datasets such as hospital morbidity.
- Through data linkage the study was able to investigate characteristics of children known to have ID by whether or not they were not identified with ID within hospital morbidity data.
- One limitation is that for some conditions associated with ID and used to identify ID in hospital codes, not all children will necessarily meet criteria for intellectual disability.
- The ICD9/10 coding system is limited in its provision of delineation of some genetic syndromes, however the integration of Orphanet coding into ICD-11 will allow many more genetic ID syndromes to be specifically identified in hospital morbidity data.

Introduction

Intellectual disability (ID) is characterised by globally impaired cognitive functioning and significant deficits in adaptive functioning, manifest before the age of 18 years.[1] Comorbid medical or psychiatric conditions are common in people with ID,[2] leading to increased hospitalisations. The increased risk of admission has been shown to range from two-fold for those with ID associated with autism up to ten-fold for those with severe ID.[3] For conditions typically managed through ambulatory (out-patient) care, people with ID have been shown to have a six-fold increase in risk of hospitalisations compared to those without ID.[4] Epilepsy is one of the most common health conditions in this population with a prevalence of around 20%[2, 5] and is one of the main reasons for hospital admission.[4] Specific disorders consistent with ID such as Down syndrome are often associated with multiple medical conditions (e.g. cardiac defects, ear disease, respiratory infections) which often require hospitalisation.[6] Mental health disorders are also more prevalent in individuals with ID [7] and hospitalisation is common.[4, 8]

Children and young adults with ID however, form a heterogeneous group and reliable population-based cohorts are not often available. Researchers investigating ID may use health data as well as other administrative datasets relating to education or service provision as well as household surveys as their sampling strategy.[9] Studies have also used health related datasets including insurance claims to identify ID and investigate specific causes of hospitalisation in this population.[4, 10]

In Western Australia the IDEA database is a population-based register of children with ID, with ascertainment from both disability service providers and education sources.[11] It is a research infrastructure that can be linked to other population datasets such as hospital morbidity data.[12] The current study aims to investigate how well the Western Australian Hospital Morbidity Data System (HMDS), which contains all admissions to private and public hospitals, recorded ID using the designated ICD codes compared to the IDEA database and thus assess the usefulness of hospitalisation data as a source of ID status.

Methods

The study cohort was restricted to children and young adults born between 1983 and 2010 and who were identified with ID in either the HMDS or the IDEA database over this period. Individuals were defined as having an intellectual disability in the HMDS if they were assigned any of the following International Classification of Diseases (ICD) diagnostic codes during hospitalisation: Mental retardation (ICD-9-CM 317-319; ICD-10-AM F70-F79), Down syndrome [Trisomy 21] (ICD-9-CM 758.0; ICD-10-AM Q90.0-Q90.2, Q90.9), Edwards/Patau syndrome [Trisomy 18/13] (ICD-9-CM 758.1, 758.2; ICD-10-AM Q91.0-Q91.7), Trisomy 9/8 (ICD-9-CM 758.5; ICD-10-AM Q92.0-Q92.5), Chromosomal deletions (ICD-9-CM 758.3; ICD-10-AM Q93.3-Q93.5), Fragile X syndrome (ICD-9-CM 759.83; ICD-10-AM Q99.2), Neurofibromatosis (ICD-9-CM 237.7; ICD-10-AM Q85.0), Tuberous sclerosis (ICD-9-CM 759.5; ICD-10-AM Q85.1), Prader-Willi syndrome (ICD-9-CM 759.81; ICD-10-AM Q87.14), and Marfan syndrome (ICD-9-CM 759.82; ICD-10-AM Q87.4). Individuals diagnosed with an intellectual disability in the IDEA database, considered the “gold standard” for ID

diagnosis in the Western Australian population, have a confirmed IQ<70 with adaptive behaviour deficits. The IDEA database and the HMDS data were linked to investigate the proportion of children confirmed with ID through IDEA who were also identified as having an ID from any one of their HMDS ICD codes. Maternal race (Aboriginal, non-Aboriginal), gender (male or female) and date of birth were obtained by linkage to the Midwives' Notification System. Information on deaths was obtained by linkage to the WA Mortality database and children and those who had died before one year of age were identified.

Age at admission (<1, 1-2, 3-5, 6-12 and >12 years), gender (male, female), race (non-Aboriginal, Aboriginal) and level of ID (mild or moderate, severe) of individuals with an intellectual disability in the IDEA dataset were compared between those who were and were not identified in the HMDS. The main cause of intellectual disability was determined through medical information recorded in the IDEA database using the Heber codes.[13] Cases with no information on cause of ID were assigned as "Unassessed". The main cause was further grouped into broad categories based on biomedical or other causes[14] in order to investigate whether the cause of ID differed between those identified and not identified with ID from the ICD codes in the HMDS dataset. Categorical variables were reported as proportions and compared using the Pearson's chi-squared test for independence. Analyses were performed using Stata 13.1 (StataCorp, College Station, Texas, USA).

The study was reviewed and approved by the Government of Western Australia Department of Health, Human Ethics Review Committee (project #2011/64).

Results

A total of 1,548,478 records representing admissions for 488,905 individuals were identified. Among them, 10,218 (2.1%) were identified as having an ID and 478,687 (97.9%) cases as not having ID in either the HMDS or the IDEA database. Those children known to IDEA who were hospitalised (n=9740), represented 92% of all children with an ID in the IDEA database (9740/10593). Of those who were diagnosed with ID, 1,435 (14.0%) were identified in both, 8,305 (81.3%) were unique to the IDEA database and 478 (4.7%) were unique to the HMDS dataset (Figure 1). Of all children identified in the HMDS dataset through the ICD codes (n=1913), 75% (n=1435) had their ID confirmed through IDEA. Death before the age of one year had occurred in 160 / 10,218 (1.5%) of the individuals identified with ID in either source with the majority (n=124, 78%) of these unique to HMDS. Limited to those who survived past one year of age, the sensitivity of using HMDS to identify ID was 14.6%, whilst the specificity was much higher at 99.9%. The positive and negative predictive values were 79.9% and 98.3% respectively.

We compared the characteristics of the 9,704 individuals who were registered in the IDEA database and thus known to have an ID, survived past one year of age and were admitted to hospital by whether they were identified with ID from the ICD codes in HMDS (Table 1).

Table 1: Characteristics of children born between 1983 and 2010 in Western Australia and survived past one year of age, who were identified with intellectual disability (ID) through the IDEA database and admitted to hospital, according to their ID diagnosis status in the Hospital Morbidity Data System (HMDS) database

Characteristic	ID diagnosis status in HMDS, N (%) Yes	No	Total	P-value*
Age at first admission (years)				
<1	1,119 (79.2)	5,636 (68.0)	6,755 (69.6)	
1-2	177 (12.3)	1,256 (15.1)	1,433 (14.7)	
3-5	54 (3.8)	714 (8.6)	768 (7.9)	<0.01
6-12	36 (2.6)	436 (5.3)	472 (4.9)	
>12	26 (1.8)	250 (3.0)	276 (2.8)	
Gender				
Male	782 (55.4)	5,489 (66.2)	6,271 (64.6)	<0.01
Female	630 (44.6)	2,803 (33.8)	3,433 (35.4)	
Race				
Non-Aboriginal	1,302 (92.2)	7,106 (85.7)	8,408 (86.6)	<0.01
Aboriginal	110 (7.8)	1,186 (14.3)	1,296 (13.4)	
Level of ID				
Mild or moderate ID	1,107 (78.4)	7,776 (93.8)	8,883 (91.5)	<0.01
Severe ID	305 (21.6)	516 (6.2)	821 (8.5)	
Total	1,412 (100)	8,292 (100)	9,704 (100)	

* Pearson's chi-squared test for independence
HMDS, Hospital Morbidity Data System; ID, intellectual disability

Children with ID who were also coded with ID in the HMDS data were more likely to be less than one year of age at first admission compared with children with ID not coded in the HMDS data (79.2% vs 68.0%). They were also more likely to be female (44.6% vs 33.8%), be non-Aboriginal (92.2% vs 85.7%) and have a severe level of ID (21.6% vs 6.2%).

Children in the IDEA database with a biomedical cause of their ID were more likely to have also been coded with ID in the HMDS dataset (Table 2).

Table 2: Cause of intellectual disability as determined in the IDEA database for children who survived to one year of age and were either identified/ not identified with ID through hospital morbidity data system (HMDS) codes

Cause of ID	In IDEA and identified with ID in HMDS		In IDEA and not identified with ID in HMDS		Total
1. PRENATAL CONDITIONS	n	%	n	%	n
Genetic or Chromosomal:					
Down Syndrome	589	94.2	36	5.8	625
Tuberous Sclerosis	29	90.6	3	9.4	32
Prader Willi Syndrome	20	87.0	3	13.0	23

Neurofibromatosis	12	70.6	5	29.4	17
Muscular Dystrophy	4	57.1	3	42.9	7
Fragile X	16	51.6	15	48.4	31
Other Chromosomal	59	45.0	72	55.0	131
X-linked inheritance	4	36.4	7	63.6	11
Metabolic (possible)	9	29.0	22	71.0	31
Myotonic Dystrophy	3	27.3	8	72.7	11
Syndrome Grouped	45	26.5	125	73.5	170
Mucopolysaccharidosis	1	25.0	3	75.0	4
Autosomal	21	23.9	67	76.1	88
Prenatal aetiology	8	18.2	36	81.8	44
Williams syndrome	5	16.1	26	83.9	31
Neurodegenerative disorders	1	11.1	8	88.9	9
Sex Chromosome	2	9.5	19	90.5	21
Mitochondria	1	7.7	12	92.3	13
Metabolic	1	5.9	16	94.1	17
Teratogenic:					
Cytomegalic Inclusion congenital	12	50.0	12	50.0	24
Other potential teratogens	4	16.7	20	83.3	24
Other prenatal infections	1	9.1	10	90.9	11
Potential Foetal alcohol syndrome	7	8.0	81	92.1	88
CNS and Other Birth Defects:					
Unspecified Neurological	32	42.7	43	57.3	75
Congenital hypothyroidism	1	25.0	3	75.0	4
Spina Bifida Meningocoele	3	25.0	9	75.0	12
Unknown Prenatal	51	22.6	175	77.4	226
Microcephaly	7	17.5	33	82.5	40
CNS: Malformations of Gyri	4	17.4	19	82.6	23
Hydrocephalus	4	16.7	20	83.3	24
Macrocephaly	3	16.7	15	83.3	18
Cranial anomalies	6	16.2	31	83.8	37
CNS Malformations	6	10.2	53	89.8	59
2. PERINATAL CONDITIONS					
Hypoxic Ischaemic Encephalopathy	27	29.0	66	71.0	93
Perinatal: Neonatal	2	28.6	5	71.4	7
3. POSTNEONATAL CONDITIONS					
Post Natal Asphyxia	13	44.8	16	55.2	29
Postnatal Injury	23	31.5	50	68.5	73
Postneonatal infection	21	29.6	50	70.4	71
Intracranial Neoplasm	2	28.6	5	71.4	7
4. NO DEFINED CAUSE					
Associated with Epilepsy	44	24.2	138	75.8	182
Cultural Familial IH	29	20.4	113	79.6	142
Associated with Coexisting disability	2	20.0	8	80.0	10
Associated with Psychotic Disorder	4	14.3	24	85.7	28
Associated Maternal medical	4	10.0	36	90.0	40

condition					
No defined cause *(Functional reaction alone)*	66	8.7	689	91.3	755
Other Developmental Disorders	3	8.3	33	91.7	36
Familial Unspecified	20	6.3	300	93.8	320
Associated with Psychosocial factors	2	6.3	30	93.8	32
Prematurity	9	6.3	133	93.7	142
Multiple Birth	2	5.0	38	95.0	40
Aspergers	1	3.9	25	96.2	26
Autism	42	3.0	1,342	97.0	1,384
Intrauterine growth restriction	1	2.9	34	97.1	35
Unassessed	114	2.7	4,103	97.3	4,217
Total	1,412	14.6	8,292	85.4	9,704

The causes most likely to have been identified with ID were Down syndrome (94.2%), Tuberous sclerosis (90.6%), Prader-Willi syndrome (87.0%), Neurofibromatosis (70.6%), muscular dystrophy (57.1%) and Fragile X (51.6%). Those least likely to have been identified with ID were those with an unassessed cause (2.7%), autism (3.0%) Asperger's (3.9%), foetal alcohol syndrome (8.0%) and other associated conditions such as intrauterine growth restriction (2.9%) and prematurity (5.6%) (Table 2). Additionally, 30% of children who had been identified with any epilepsy diagnosis in the IDEA database, regardless of their main cause of ID diagnosis, were found to be identified with ID in the hospital dataset (not shown in Table). Of the children who were identified through both IDEA and HMDS and survived one year of age (n=1412), 42% had been identified in the IDEA database as having Down syndrome, a condition representing about 7% of all those with ID or about 10% of those where information on cause of ID may be available through disability services.

Children identified with ID in the HMDS dataset who were not in the IDEA database and had survived one year were investigated according to the ICD codes used to identify ID in HMDS (Table 3).

Table 3: Children born between 1983 and 2010 in Western Australia and were identified with intellectual disability (ID) through ICD codes in the Hospital Morbidity Data System (HMDS) database but not identified in the IDEA database, by death status and ID diagnosis in HMDS

ID diagnosis in HMDS	Died under one year	Alive after one year	Total
	n (%)	n (%)	n (%)
Mental retardation	3 (2.4)	138 (39.0)	141 (29.5)
Down syndrome	25 (20.2)	45 (12.7)	70 (14.6)
Trisomy 18/13	80 (64.5)	5 (1.4)	85 (17.8)
Trisomy 8/9	10 (8.1)	12 (3.4)	22 (4.6)
Chromosomal deletion	5 (4.0)	16 (4.5)	21 (4.4)
Fragile X	0	1 (0.3)	1 (0.2)
Neurofibromatosis	0	79 (22.3)	79 (16.5)
Tuberous sclerosis	0	17 (4.8)	17 (3.6)
Prader-Willi syndrome	0	6 (1.7)	6 (1.3)
Marfan syndrome	1 (0.8)	35 (9.9)	36 (7.5)
Total	124	354	478

The majority of those not in IDEA had been assigned an ICD code aligned to mental retardation (n=138, 39.0%), Neurofibromatosis (n=79, 22.3%) or Down syndrome (n=45, 12.7%) (Table 3). Among the 124 (25.9%) individuals who had died before one year of age, 75% had died before one month, and the majority of diagnoses included Trisomy 18/13 (n=80, 64.5%), Down syndrome (n=25, 20.2%) or Trisomy 8/9 (n=10, 8.1%). If it is assumed that all additional cases identified through ICD codes but not in the IDEA database did have ID (n=478), then the completeness of ascertainment in IDEA would represent 95.7%. With the assumption that those who died under one year would not be able to be ascertained (n=124, of whom the majority died under one month) then IDEA would represent 96.8%.

Discussion

Data from Western Australia suggest that hospital morbidity data may be an inadequate source of identification of intellectual disability in epidemiological studies with a sensitivity of only 14%. After removing children who died before one year of age, intellectual disability of syndromic or monogenic aetiology such as that associated with Down syndrome, Neurofibromatosis and Fragile X syndrome was most likely also to be identified in hospital sources and ID of unknown cause least likely to be identified. Females and children under one year were also more likely to be identified while Aboriginal children and those with a mild-moderate level of intellectual disability were less likely to be identified.

The greatest strength of this study was the availability of a population source of ID, the IDEA database which has used both disability service use and education sources to maintain high ascertainment over the last thirty years.[15] It has already been used as a data source for multiple data linkage studies investigating determinants[16-18] and outcomes[3, 19] associated with intellectual disability. One limitation is the lack of information on cause of ID for those cases ascertained only through education sources, as medical information is obtained through the referral process to disability services. Another limitation is that there are several conditions where only a percentage of children have an intellectual disability, in contrast to conditions like Down syndrome where almost all children are affected. However for the purposes of this study we still elected to use the ICD codes for these diagnoses to identify ID in the HMDS in order to capture the maximum possible number of children with ID. Thus by doing this and assigning ID status to all children with these conditions in

hospital morbidity records we could have overestimated the number with ID. For example, intellectual disability is diagnosed in approximately half of individuals with tuberous sclerosis[20] and whilst almost all of those with Prader Willi syndrome will have cognitive deficits, up to 40% may fall within the borderline range.[21] About a third of children with neurofibromatosis have been reported to have general learning difficulties associated with borderline or lower IQ[22] and children with Marfan syndrome may only have a slightly increased risk of intellectual disability.[23] Children diagnosed with autism spectrum disorder have been found to have an ID in approximately 30%- 60% of cases although this proportion has been shown to be decreasing in more recent years.[17, 24, 25] The effect of removing these conditions from our HMDS search list would have been to slightly increase the sensitivity and positive predictive value of using HMDS to identify ID.

Children with a cause of ID commonly known to be associated with ID, such as Down syndrome or Prader Willi syndrome, were most likely to be identified with ID in the hospital data, possibly due to the fact that these codes had been specifically designated in the ICD search codes for ID, unlike those for whom no clear cause had been recorded in the IDEA Database. The inability of ICD codes to specifically identify relatively rare conditions is also problematic if relying on such identification of ID. For example, Williams syndrome, known to be highly associated with ID,[26] is identified with a Q89.8 ICD-10 code which is in itself not specific for Williams syndrome and was not used in our search strategy as it would also identify children possibly without ID such as those with Stickler syndrome. Perhaps as a consequence, children with Williams syndrome were poorly identified as ID in the hospital codes,

with only 16% of children being coded as such. Recent versions of ICD-10-AM provide a finer delineation of genetic syndromes and thus allow better differentiation of syndromes with ID from those without the condition. The integration of Orphanet coding into ICD-11 will allow many more genetic ID syndromes to be specifically identified in hospital morbidity data.[27] This has become a matter of urgency given the accelerated identification of these genetic causes over the last decade and particularly since the introduction of next generation sequencing.[28-30]

Many children who would be expected to develop ID by virtue of their diagnosis experience serious and life-threatening comorbidities and as a consequence may die early. As we have shown, about a third of those not identified in the IDEA database had died, nearly three quarters before one month of age and the majority by one year. In these cases it would be unlikely that families would have sought registration for disability services before their child died and hence they would not have been included within the IDEA database. The remaining cases identified with ID through the hospital ICD codes but who were not in IDEA represent potential missed ascertainment within IDEA, however this number is relatively small, effectively reducing the completeness of IDEA to 96% if these cases had met eligibility for inclusion in IDEA. There is the possibility that some of these, most likely those with neurofibromatosis, tuberous sclerosis, Marfan syndrome or Prader Willi syndrome may have a milder cognitive deficit and not meet the criteria for ID.

We found one Canadian study which had used hospital morbidity codes to identify ID in at least one patient record in order to form their cohort, but had found that as many as half of the multiple records for these individuals did not code ID as a comorbidity in the hospital morbidity system.[8] It was therefore likely that other individuals with ID had been missed from their cohort due to inconsistent coding of ID as a comorbidity. The authors acknowledged that, similar to our own findings, it was likely that those who had been identified with ID were more severe. In a later study they estimated population prevalence of ID by identifying individuals with ID using the same ICD codes related to ID in a number of different health administrative datasets.[31] Using their broadest capture algorithm they found an overall prevalence of 8/1000 and a prevalence of 14.2/1000 in young adults aged 18-24 years,[31] not too dissimilar from our own estimate of 17/1000 in a similarly aged population.[15] These prevalence estimates based on health datasets certainly provided better ascertainment than the 14% capture using our own hospital morbidity codes but the ascertainment is still likely less complete than our population ascertainment using the IDEA database. Linked data studies in New South Wales, Australia have provided further evidence of the need for multiple sources of ascertainment of ID.[32] Using ICD codes for ID within health datasets, as well as disability services, birth and mortality linkages, the authors found an overall prevalence of 0.6% (or 6/1000) considerably less than that found in our IDEA database.

Practical considerations for clinical care would suggest that hospital coding which does not include reference to intellectual disability as a comorbidity may impact on the way in which service is delivered to this particularly vulnerable population.

Better coding practices for ID would enable researchers to investigate directly whether care or procedures are compromised for individuals with ID and facilitate the development of ID-related policies and service planning. The hospital experiences for people with ID, who we know experience higher rates of hospitalisation than the rest of the population,[3] have been described as relying heavily on carers for in-hospital patient assistance with failure to provide appropriate care, and lack of knowledge and discharge planning by medical staff.[33] The reliance on hospital morbidity data, as well as other administrative datasets, to identify ID in a population for research purposes has been shown to provide varied results.[9] Overall, we would not recommend that researchers use hospital morbidity datasets alone as a source of identification of intellectual disability.

Conclusion

Through linkage to a hospital morbidity dataset, this study has shown that hospital data does not adequately identify individuals with ID when compared with the population-based IDEA database. A high proportion of those uniquely identified in hospital morbidity data had died early or alternatively they had a condition not necessarily associated with ID. It is important for hospital codes to reflect the ID status of patients, primarily for the benefit of recognizing their specific needs, but also for improvement of ascertainment of ID through this source. Clearly with such a high proportion of individuals not being recognized with ID, coding practices which identify ID need to be better implemented.

Acknowledgements

The authors acknowledge the Disability Services Commission, the Telethon Kids Institute, the Western Australia Department of Education, the Catholic Education Office and the Association of Independent Schools of Western Australia for assistance with data collection for the IDEA database.

Funding Statement

This study was partly funded by an Australian National Health and Medical Research Council (NHMRC) Program Grant #572742.

Competing Interests

We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Author Contributions

All authors contributed to the initial design of the manuscript. JB and HL were responsible for the drafting of the paper. KW was responsible for analysis and contributed to the writing of the final draft. All authors contributed to the final writing of the paper and checked for important intellectual content.

Data Sharing Statement

Data are only available through ethical approval from the Western Australian Department of Health, Human Ethics Review Committee in collaboration with the authors.

References

1. Schalock R, Borthwick-Duffy S, Bradley V, Buntinx W, Coulter D, Craig E, et al. Intellectual disability: Definition, classification, and systems of supports: American Association on Intellectual and Developmental Disabilities. 444 North Capitol Street NW Suite 846, Washington, DC 20001; 2010.

2. Oeseburg B, Dijkstra GJ, Groothoff JW, Reijneveld SA, Jansen DE. Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review. Intellect Dev Disabil. 2011;49:59-85.

3. Bebbington A, Glasson E, Bourke J, de Klerk N, Leonard H. Hospitalisation rates for children with intellectual disability or autism born in Western Australia 1983-1999: a population-based cohort study. BMJ open. 2013;3.

4. Balogh R, Brownell M, Ouellette-Kuntz H, Colantonio A. Hospitalisation rates for ambulatory care sensitive conditions for persons with and without an intellectual disability--a population perspective. J Intellect Disabil Res. 2010;54:820-32.

5. Bowley C, Kerr M. Epilepsy and intellectual disability. J Intellect Disabil Res. 2000;44 (Pt 5):529-43.

6. Fitzgerald P, Leonard H, Pikora TJ, Bourke J, Hammond G. Hospital admissions in children with Down syndrome: experience of a population-based cohort followed from birth. PloS one. 2013;8:e70401.

7. Emerson E, Hatton C. Mental health of children and adolescents with intellectual disabilities in Britain. Br J Psychiatry. 2007;191:493-9.

8. Balogh RS, Hunter D, Ouellette-Kuntz H. Hospital Utilization among Persons with an Intellectual Disability, Ontario, Canada, 1995–2001. J Appl Res Intellect. 2005;18:181-90.

9. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. Res Dev Disabil. 2011;32:419-36.

10. Büchele G, Becker C, Cameron I, Auer R, Rothenbacher D, König H, et al. Fracture risk in people with developmental disabilities: results of a large claims data analysis. Osteoporosis Int. 2017;28:369-75.

11. Petterson B, Leonard H, Bourke J, Sanders R, Chalmers R, Jacoby P, et al. IDEA (Intellectual Disability Exploring Answers): a population-based database for intellectual disability in Western Australia. Ann Hum Biol. 2005;32:237-43.

12. Holman CDAJ, Bass AJ, Rouse IL, Hobbs MST. Population-based linkage of health records in Western Australia: development of a health services research linked database. Aust NZ J Publ Heal. 1999;23:453-9.

13. Heber R. A manual on terminology and classification in mental retardation. Am J Ment Def. 1961;64.

14. Yeargin‐Allsopp M, Murphy CC, Cordrro JF, Decoufle P, Hollowell JG. Reported biomedical causes and associated medical conditions for mental retardation among 10‐year‐old children, metropolitan Atlanta, 1985 to 1987. Dev Med Child Neurol. 1997;39:142-9.

15. Bourke J, de Klerk N, Smith T, Leonard H. Population-Based Prevalence of Intellectual Disability and Autism Spectrum Disorders in Western Australia: A Comparison With Previous Estimates. Medicine. 2016;95:e3737.

16. Leonard H, Glasson E, Bebbington A, Hammond G, Croft D, Pikora T, et al. Application of Population-Based Linked Data to the Study of Intellectual Disability and Autism. In: Urbano RC, editor. International Review of Research in Developmental Disabilities. 45. MA, USA: Academic Press; 2013. p. 281-327.

17. Leonard H, Glasson E, Nassar N, Whitehouse A, Bebbington A, Bourke J, et al. Autism and intellectual disability are differentially related to sociodemographic background at birth. PloS one. 2011;6:e17875.

18. Leonard H, Nassar N, Bourke J, Blair E, Mulroy S, deKlerk N, et al. Intrauterine growth and subsequent intellectual disability in a ten-year population cohort of children in Western Australia. Am J Epidemiol. 2007;165:S19-S.

19. Bourke J, Nembhard WN, Wong K, Leonard H. Twenty-Five Year Survival of Children with Intellectual Disability in Western Australia. J Pediatr. 2017.

20. Curatolo P, Bombardieri R, Jozwiak S. Tuberous sclerosis. The Lancet. 2008;372:657-68.

21. Cassidy SB, Driscoll DJ. Prader–Willi syndrome. Eur J Hum Genet. 2009;17:3-13.

22. Hyman SL, Arthur E, North KN. Learning disabilities in children with neurofibromatosis type 1: subtypes, cognitive profile, and attention-deficit-hyperactivity disorder. Dev Med Child Neurol. 2006;48:973-7.

23. Dietz H. Marfan Syndrome Seattle (WA): University of Washington, Seattle; 2001 [updated 2 Feb, 201711 July, 2017]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1335/.

24. Baio J. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. MMWR Surveill Summ. 2014;63:1-21.

25. Nassar N, Dixon G, Bourke J, Bower C, Glasson E, de Klerk N, et al. Autism spectrum disorders in young children: effect of changes in diagnostic practices. Int J Epidemiol. 2009;38:1245-54.

26. Greer MK, Brown FR, Pai GS, Choudry SH, Klein AJ. Cognitive, adaptive, and behavioral characteristics of Williams syndrome. Am J Med Genet. 1997;74:521-5.

27. Walker CE, Mahede T, Davis G, Miller LJ, Girschik J, Brameld K, et al. The collective impact of rare diseases in Western Australia: an estimate using a population-based cohort. Genet Med. 2017;19:546-52.

28. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511:344-+.

29. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2016;17:9-18.

30. Valdez R, Grosse SD, Khoury MJ. The need for a next-generation public health response to rare diseases. Genet Med. 2016.

31. Lin E, Balogh R, Cobigo V, Ouellette-Kuntz H, Wilton AS, Lunsky Y. Using administrative health data to identify individuals with intellectual and developmental disabilities: a comparison of algorithms. J Intellect Disabil Res. 2013;57:462-77.

32. Reppermund S, Srasuebkul P, Heintze T, Reeve R, Dean K, Emerson E, et al. Cohort profile: a data linkage cohort to examine health service profiles of people with intellectual disability in New South Wales, Australia. BMJ open. 2017;7:e015627.

33. Iacono T, Bigby C, Unsworth C, Douglas J, Fitzpatrick P. A systematic review of hospital experiences of people with intellectual disability. BMC Health Serv Res. 2014;14:505.

Figure Legend:
Figure 1: Identification of intellectual disability (ID) in children born 1983-2010 and hospitalised in Western Australia using linkage to the IDEA database and the hospital morbidity data system (HMDS)

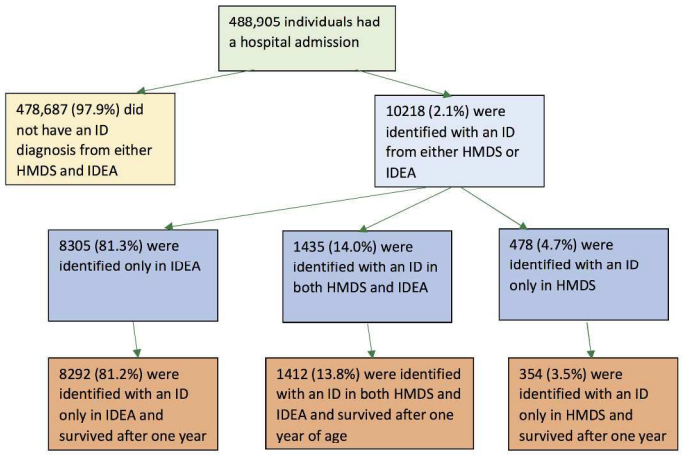

Figure 1: Identification of intellectual disability (ID) in children born 1983-2010 and hospitalised in Western Australia using linkage to the IDEA database and the hospital morbidity data system (HMDS)

209x296mm (300 x 300 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 5
Methods			
Study design	4	Present key elements of study design early in the paper	Pages 5,6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Pages 5,6
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Page 6
Bias	9	Describe any efforts to address potential sources of bias	n/a
Study size	10	Explain how the study size was arrived at	Page 5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Page 6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 6
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6
		(*c*) Explain how missing data were addressed	n/a
		(*d*) If applicable, explain how loss to follow-up was addressed	n/a
		(*e*) Describe any sensitivity analyses	n/a
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Page 7
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) Summarise follow-up time (eg, average and total amount)	n/a
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Page 7
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n/a
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 8,9
Limitations			Page 9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Pages 10,11
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Validation of intellectual disability coding through hospital morbidity records using an intellectual disability population-based database in Western Australia

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019113.R2
Article Type:	Research
Date Submitted by the Author:	04-Dec-2017
Complete List of Authors:	Bourke, Jenny; Telethon Kids Institute, Epidemiology Wong, Kingsley; Telethon Kids Institute Leonard, Helen
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Research methods
Keywords:	

Validation of intellectual disability coding through hospital morbidity records using an intellectual disability population-based database in Western Australia

Jenny Bourke, MPH[1] Kingsley Wong, MBBS[1] Helen Leonard, MBChB[1]

[1] Telethon Kids Institute, University of Western Australia, Perth, Western Australia, Australia

Corresponding author

Helen Leonard

helen.leonard@telethonkids.org.au

MOB:+61 419956946

Telethon Kids Institute

100 Roberts Rd, Subiaco

WA 6008 Australia

Word count=2580

Abstract

Objectives: To investigate how well intellectual disability can be ascertained using hospital morbidity data compared with a population-based data source.

Design, Setting and Participants: All children born 1983-2010 with a hospital admission in the Western Australian Hospital Morbidity Data System (HMDS) were linked with the Western Australian Intellectual Disability (IDEA) database. The ICD hospital codes consistent with intellectual disability were also identified.

Main Outcome measures: The characteristics of those children identified with intellectual disability through either or both sources were investigated.

Results: Of the 488,905 individuals in the study, 10,218 (2.1%) were identified with intellectual disability in either IDEA or HMDS with 1,435 (14.0%) individuals identified in both databases, 8,305 (81.3%) unique to the IDEA database and 478 (4.7%) unique to the HMDS dataset only. Of those unique to the HMDS dataset, about a quarter (n=124) had died before one year of age and most of these (75%) before one month. Children with intellectual disability who were also coded as such in the HMDS data were more likely to be aged under one year, female, non-Aboriginal and have a severe level of intellectual disability, compared with those not coded in the HMDS data. The sensitivity of using HMDS to identify intellectual disability was 14.7%, whilst the specificity was much higher at 99.9%.

Conclusion: Hospital morbidity data are not a reliable source for identifying intellectual disability within a population and epidemiological researchers need to take these findings into account in their study design.

Strengths and limitations of this study

- The greatest strength of this study was the availability of a population-based source of intellectual disability (ID).
- The state-wide data linkage system allowed this database to be linked to other population datasets such as hospital morbidity.
- Through data linkage the study was able to investigate characteristics of children known to have ID by whether or not they were not identified with ID within hospital morbidity data.
- One limitation is that for some conditions associated with ID and used to identify ID in hospital codes, not all children will necessarily meet criteria for intellectual disability.
- The ICD9/10 coding system is limited in its provision of delineation of some genetic syndromes, however the integration of Orphanet coding into ICD-11 will allow many more genetic ID syndromes to be specifically identified in hospital morbidity data.

Introduction

Intellectual disability (ID) is characterised by globally impaired cognitive functioning and significant deficits in adaptive functioning, manifest before the age of 18 years.[1] Comorbid medical or psychiatric conditions are common in people with ID,[2] leading to increased hospitalisations. The increased risk of admission has been shown to range from two-fold for those with ID associated with autism up to ten-fold for those with severe ID.[3] For conditions typically managed through ambulatory (out-patient) care, people with ID have been shown to have a six-fold increase in risk of hospitalisations compared to those without ID.[4] Epilepsy is one of the most common health conditions in this population with a prevalence of around 20%[2, 5] and is one of the main reasons for hospital admission.[4] Specific disorders consistent with ID such as Down syndrome are often associated with multiple medical conditions (e.g. cardiac defects, ear disease, respiratory infections) which often require hospitalisation.[6] Mental health disorders are also more prevalent in individuals with ID [7] and hospitalisation is common.[4, 8]

Children and young adults with ID however, form a heterogeneous group and reliable population-based cohorts are not often available. Researchers investigating ID may use health data as well as other administrative datasets relating to education or service provision as well as household surveys as their sampling strategy.[9] Studies have also used health related datasets including insurance claims to identify ID and investigate specific causes of hospitalisation in this population.[4, 10]

In Western Australia the IDEA database is a population-based register of children with ID, with ascertainment from both disability service providers and education sources.[11] It is a research infrastructure that can be linked to other population datasets such as hospital morbidity data.[12] The current study aims to investigate how well the Western Australian Hospital Morbidity Data System (HMDS), which contains all admissions to private and public hospitals, recorded ID using the designated ICD codes compared to the IDEA database and thus assess the usefulness of hospitalisation data as a source of ID status.

Methods

The study cohort was restricted to children and young adults born between 1983 and 2010 and who were identified with ID in either the HMDS or the IDEA database over this period. Individuals were defined as having an intellectual disability in the HMDS if they were assigned any of the following International Classification of Diseases (ICD) diagnostic codes during hospitalisation: Mental retardation (ICD-9-CM 317-319; ICD-10-AM F70-F79), Down syndrome [Trisomy 21] (ICD-9-CM 758.0; ICD-10-AM Q90.0-Q90.2, Q90.9), Edwards/Patau syndrome [Trisomy 18/13] (ICD-9-CM 758.1, 758.2; ICD-10-AM Q91.0-Q91.7), Trisomy 9/8 (ICD-9-CM 758.5; ICD-10-AM Q92.0-Q92.5), Chromosomal deletions (ICD-9-CM 758.3; ICD-10-AM Q93.3-Q93.5), Fragile X syndrome (ICD-9-CM 759.83; ICD-10-AM Q99.2), Neurofibromatosis (ICD-9-CM 237.7; ICD-10-AM Q85.0), Tuberous sclerosis (ICD-9-CM 759.5; ICD-10-AM Q85.1), Prader-Willi syndrome (ICD-9-CM 759.81; ICD-10-AM Q87.14), and Marfan syndrome (ICD-9-CM 759.82; ICD-10-AM Q87.4). ICD coding in the hospital morbidity dataset is completed by clinical coders who abstract relevant information from the

patient's medical record and decide which diagnoses and procedures meet criteria for coding as per Australian and WA Coding Standards.

Individuals diagnosed with an intellectual disability in the IDEA database, considered the "gold standard" for ID diagnosis in the Western Australian population, have a confirmed IQ<70 with adaptive behaviour deficits. The IDEA database and the HMDS data were linked to investigate the proportion of children confirmed with ID through IDEA who were also identified as having an ID from any one of their HMDS ICD codes. Maternal race (Aboriginal, non-Aboriginal), gender (male or female) and date of birth were obtained by linkage to the Midwives' Notification System. Information on deaths was obtained by linkage to the WA Mortality database and children and those who had died before one year of age were identified.

Age at admission (<1, 1-2, 3-5, 6-12 and >12 years), gender (male, female), race (non-Aboriginal, Aboriginal) and level of ID (mild or moderate, severe) of individuals with an intellectual disability in the IDEA dataset were compared between those who were and were not identified in the HMDS. The main cause of intellectual disability was determined by medical personnel at the Disability Services Commission from medical records and recorded in the IDEA database using the Heber codes.[13] Cases with no information on cause of ID were assigned as "Unassessed". The main cause was further grouped into broad categories based on biomedical or other causes[14] in order to investigate whether the cause of ID differed between those identified and not identified with ID from the ICD codes in the HMDS dataset. Categorical variables were reported as proportions and compared using the

Pearson's chi-squared test for independence. Analyses were performed using Stata 13.1 (StataCorp, College Station, Texas, USA).

The study was reviewed and approved by the Government of Western Australia Department of Health, Human Ethics Review Committee (project #2011/64).

Results

A total of 1,548,478 records representing admissions for 488,905 individuals were identified. Among them, 10,218 (2.1%) were identified as having an ID and 478,687 (97.9%) cases as not having ID in either the HMDS or the IDEA database. Those children known to IDEA who were hospitalised (n=9740), represented 92% of all children with an ID in the IDEA database (9740/10593). Of those who were diagnosed with ID, 1,435 (14.0%) were identified in both, 8,305 (81.3%) were unique to the IDEA database and 478 (4.7%) were unique to the HMDS dataset (Figure 1). Of all children identified in the HMDS dataset through the ICD codes (n=1913), 75% (n=1435) had their ID confirmed through IDEA. Death before the age of one year had occurred in 160 / 10,218 (1.5%) of the individuals identified with ID in either source with the majority (n=124, 78%) of these unique to HMDS. Limited to those who survived past one year of age, the sensitivity of using HMDS to identify ID was 14.6%, whilst the specificity was much higher at 99.9%. The positive and negative predictive values were 79.9% and 98.3% respectively.

We compared the characteristics of the 9,704 individuals who were registered in the IDEA database and thus known to have an ID, survived past one year of age and

were admitted to hospital by whether they were identified with ID from the ICD codes in HMDS (Table 1).

Table 1: Characteristics of children born between 1983 and 2010 in Western Australia and survived past one year of age, who were identified with intellectual disability (ID) through the IDEA database and admitted to hospital, according to their ID diagnosis status in the Hospital Morbidity Data System (HMDS) database

	ID diagnosis status in HMDS, N (%)			
Characteristic	**Yes**	**No**	**Total**	**P-value***
Age at first admission (years)				
<1	1,119 (79.2)	5,636 (68.0)	6,755 (69.6)	
1-2	177 (12.3)	1,256 (15.1)	1,433 (14.7)	
3-5	54 (3.8)	714 (8.6)	768 (7.9)	<0.01
6-12	36 (2.6)	436 (5.3)	472 (4.9)	
>12	26 (1.8)	250 (3.0)	276 (2.8)	
Gender				
Male	782 (55.4)	5,489 (66.2)	6,271 (64.6)	<0.01
Female	630 (44.6)	2,803 (33.8)	3,433 (35.4)	
Race				
Non-Aboriginal	1,302 (92.2)	7,106 (85.7)	8,408 (86.6)	<0.01
Aboriginal	110 (7.8)	1,186 (14.3)	1,296 (13.4)	
Level of ID				
Mild or moderate ID	1,107 (78.4)	7,776 (93.8)	8,883 (91.5)	<0.01
Severe ID	305 (21.6)	516 (6.2)	821 (8.5)	
Total	1,412 (100)	8,292 (100)	9,704 (100)	

* Pearson's chi-squared test for independence
HMDS, Hospital Morbidity Data System; ID, intellectual disability

Children with ID who were also coded with ID in the HMDS data were more likely to be less than one year of age at first admission compared with children with ID not coded in the HMDS data (79.2% vs 68.0%). They were also more likely to be female (44.6% vs 33.8%), be non-Aboriginal (92.2% vs 85.7%) and have a severe level of ID (21.6% vs 6.2%).

Children in the IDEA database with a biomedical cause of their ID were more likely to have also been coded with ID in the HMDS dataset (Table 2).

Table 2: Cause of intellectual disability as determined in the IDEA database for children who survived to one year of age and were either identified/ not identified with ID through hospital morbidity data system (HMDS) codes

Cause of ID	In IDEA and identified with ID in HMDS		In IDEA and not identified with ID in HMDS		Total
1. PRENATAL CONDITIONS	n	%	n	%	n
Genetic or Chromosomal:					
Down Syndrome	589	94.2	36	5.8	625
Tuberous Sclerosis	29	90.6	3	9.4	32
Prader Willi Syndrome	20	87.0	3	13.0	23
Neurofibromatosis	12	70.6	5	29.4	17
Muscular Dystrophy	4	57.1	3	42.9	7
Fragile X	16	51.6	15	48.4	31
Other Chromosomal	59	45.0	72	55.0	131
X-linked inheritance	4	36.4	7	63.6	11
Metabolic (possible)	9	29.0	22	71.0	31
Myotonic Dystrophy	3	27.3	8	72.7	11
Syndrome Grouped	45	26.5	125	73.5	170
Mucopolysaccharidosis	1	25.0	3	75.0	4
Autosomal	21	23.9	67	76.1	88
Prenatal aetiology	8	18.2	36	81.8	44
Williams syndrome	5	16.1	26	83.9	31
Neurodegenerative disorders	1	11.1	8	88.9	9
Sex Chromosome	2	9.5	19	90.5	21
Mitochondria	1	7.7	12	92.3	13
Metabolic	1	5.9	16	94.1	17
Teratogenic:					
Cytomegalic Inclusion congenital	12	50.0	12	50.0	24
Other potential teratogens	4	16.7	20	83.3	24
Other prenatal infections	1	9.1	10	90.9	11
Potential Foetal alcohol syndrome	7	8.0	81	92.1	88
CNS and Other Birth Defects:					
Unspecified Neurological	32	42.7	43	57.3	75
Congenital hypothyroidism	1	25.0	3	75.0	4
Spina Bifida Meningocoele	3	25.0	9	75.0	12
Unknown Prenatal	51	22.6	175	77.4	226
Microcephaly	7	17.5	33	82.5	40
CNS: Malformations of Gyri	4	17.4	19	82.6	23
Hydrocephalus	4	16.7	20	83.3	24
Macrocephaly	3	16.7	15	83.3	18
Cranial anomalies	6	16.2	31	83.8	37
CNS Malformations	6	10.2	53	89.8	59
2. PERINATAL CONDITIONS					
Hypoxic Ischaemic Encephalopathy	27	29.0	66	71.0	93
Perinatal: Neonatal	2	28.6	5	71.4	7

3. POSTNEONATAL CONDITIONS					
Post Natal Asphyxia	13	44.8	16	55.2	29
Postnatal Injury	23	31.5	50	68.5	73
Postneonatal infection	21	29.6	50	70.4	71
Intracranial Neoplasm	2	28.6	5	71.4	7
4. NO DEFINED CAUSE					
Associated with Epilepsy	44	24.2	138	75.8	182
Cultural Familial IH	29	20.4	113	79.6	142
Associated with Coexisting disability	2	20.0	8	80.0	10
Associated with Psychotic Disorder	4	14.3	24	85.7	28
Associated Maternal medical condition	4	10.0	36	90.0	40
No defined cause *(Functional reaction alone)*	66	8.7	689	91.3	755
Other Developmental Disorders	3	8.3	33	91.7	36
Familial Unspecified	20	6.3	300	93.8	320
Associated with Psychosocial factors	2	6.3	30	93.8	32
Prematurity	9	6.3	133	93.7	142
Multiple Birth	2	5.0	38	95.0	40
Aspergers	1	3.9	25	96.2	26
Autism	42	3.0	1,342	97.0	1,384
Intrauterine growth restriction	1	2.9	34	97.1	35
Unassessed	114	2.7	4,103	97.3	4,217
Total	1,412	14.6	8,292	85.4	9,704

The causes in IDEA most likely to have also been identified with ID in any of the HMDS ICD codes were Down syndrome (94.2%), Tuberous sclerosis (90.6%), Prader-Willi syndrome (87.0%), Neurofibromatosis (70.6%), muscular dystrophy (57.1%) and Fragile X (51.6%). Those least likely to have been identified with ID were those with an unassessed cause (2.7%), autism (3.0%) Asperger's (3.9%), foetal alcohol syndrome (8.0%) and other associated conditions such as intrauterine growth restriction (2.9%) and prematurity (5.6%) (Table 2). Additionally, 30% of children who had been identified with any epilepsy diagnosis in the IDEA database, regardless of their main cause of ID diagnosis, were found to be identified with ID in the hospital dataset (not shown in Table). For the children who were identified through both IDEA and HMDS and survived one year of age (n=1412), n=623 had an ICD code

for “mental retardation”. For the remaining n=789, the consensus of diagnosis between IDEA and the ICD codes for particular disorders was 80-98% for Down syndrome, Trisomy 18/13, Trisomy 9/8, Chromosomal deletions, Fragile X syndrome, Tuberous sclerosis and Prader-Willi syndrome; and less for Neurofibromatosis (63%) and Marfan syndrome (12.5%).

Children identified with ID in the HMDS dataset who were not in the IDEA database and had survived one year were investigated according to the ICD codes used to identify ID in HMDS (Table 3).

Table 3: Children born between 1983 and 2010 in Western Australia and were identified with intellectual disability (ID) through ICD codes in the Hospital Morbidity Data System (HMDS) database but not identified in the IDEA database, by death status and ID diagnosis in HMDS

ID diagnosis in HMDS	Died under one year	Alive after one year	Total
	n (%)	n (%)	n (%)
Mental retardation	3 (2.4)	138 (39.0)	141 (29.5)
Down syndrome	25 (20.2)	45 (12.7)	70 (14.6)
Trisomy 18/13	80 (64.5)	5 (1.4)	85 (17.8)
Trisomy 8/9	10 (8.1)	12 (3.4)	22 (4.6)
Chromosomal deletion	5 (4.0)	16 (4.5)	21 (4.4)
Fragile X	0	1 (0.3)	1 (0.2)
Neurofibromatosis	0	79 (22.3)	79 (16.5)
Tuberous sclerosis	0	17 (4.8)	17 (3.6)
Prader-Willi syndrome	0	6 (1.7)	6 (1.3)
Marfan syndrome	1 (0.8)	35 (9.9)	36 (7.5)
Total	124	354	478

The majority of those not in IDEA had been assigned an ICD code aligned to mental retardation (n=138, 39.0%), Neurofibromatosis (n=79, 22.3%) or Down syndrome (n=45, 12.7%) (Table 3). Among the 124 (25.9%) individuals who had died before one year of age, 75% had died before one month, and the majority of diagnoses included Trisomy 18/13 (n=80, 64.5%), Down syndrome (n=25, 20.2%) or Trisomy 8/9 (n=10,

8.1%). If it is assumed that all additional cases identified through ICD codes but not in the IDEA database did have ID (n=478), then the completeness of ascertainment in IDEA would represent 95.7%. With the assumption that those who died under one year would not be able to be ascertained (n=124, of whom the majority died under one month) then IDEA would represent 96.8%.

Discussion

Data from Western Australia suggest that hospital morbidity data may be an inadequate source of identification of intellectual disability in epidemiological studies with a sensitivity of only 14%. After removing children who died before one year of age, intellectual disability of syndromic or monogenic aetiology such as that associated with Down syndrome, Neurofibromatosis and Fragile X syndrome was most likely also to be identified in hospital sources and ID of unknown cause least likely to be identified. Females and children under one year were also more likely to be identified while Aboriginal children and those with a mild-moderate level of intellectual disability were less likely to be identified.

The greatest strength of this study was the availability of a population source of ID, the IDEA database which has used both disability service use and education sources to maintain high ascertainment over the last thirty years.[15] It has already been used as a data source for multiple data linkage studies investigating determinants[16-18] and outcomes[3, 19] associated with intellectual disability. One limitation is the lack of information on cause of ID for those cases ascertained only through education sources, as medical information is obtained through the referral process to disability

services. Another limitation is that there are several conditions where only a percentage of children have an intellectual disability, in contrast to conditions like Down syndrome where almost all children are affected. However for the purposes of this study we still elected to use the ICD codes for these diagnoses to identify ID in the HMDS in order to capture the maximum possible number of children with ID. Thus by doing this and assigning ID status to all children with these conditions in hospital morbidity records we could have overestimated the number with ID. For example, intellectual disability is diagnosed in approximately half of individuals with tuberous sclerosis[20] and whilst almost all of those with Prader Willi syndrome will have cognitive deficits, up to 40% may fall within the borderline range.[21] About a third of children with neurofibromatosis have been reported to have general learning difficulties associated with borderline or lower IQ[22] and children with Marfan syndrome may only have a slightly increased risk of intellectual disability.[23] Children diagnosed with autism spectrum disorder have been found to have an ID in approximately 30%- 60% of cases although this proportion has been shown to be decreasing in more recent years.[17, 24, 25] The effect of removing these conditions from our HMDS search list would have been to slightly increase the sensitivity and positive predictive value of using HMDS to identify ID.

Children with a cause of ID commonly known to be associated with ID, such as Down syndrome or Prader Willi syndrome, were most likely to be identified with ID in the hospital data, possibly due to the fact that these codes had been specifically designated in the ICD search codes for ID, unlike those for whom no clear cause had been recorded in the IDEA Database. The inability of ICD codes to specifically identify

relatively rare conditions is also problematic if relying on such identification of ID. For example, Williams syndrome, known to be highly associated with ID,[26] is identified with a Q89.8 ICD-10 code which is in itself not specific for Williams syndrome and was not used in our search strategy as it would also identify children possibly without ID such as those with Stickler syndrome. Perhaps as a consequence, children with Williams syndrome were poorly identified as ID in the hospital codes, with only 16% of children being coded as such. Recent versions of ICD-10-AM provide a finer delineation of genetic syndromes and thus allow better differentiation of syndromes with ID from those without the condition. The integration of Orphanet coding into ICD-11 will allow many more genetic ID syndromes to be specifically identified in hospital morbidity data.[27] This has become a matter of urgency given the accelerated identification of these genetic causes over the last decade and particularly since the introduction of next generation sequencing.[28-30]

Many children who would be expected to develop ID by virtue of their diagnosis experience serious and life-threatening comorbidities and as a consequence may die early. As we have shown, about a third of those not identified in the IDEA database had died, nearly three quarters before one month of age and the majority by one year. In these cases it would be unlikely that families would have sought registration for disability services before their child died and hence they would not have been included within the IDEA database. The remaining cases identified with ID through the hospital ICD codes but who were not in IDEA represent potential missed ascertainment within IDEA, however this number is relatively small, effectively

reducing the completeness of IDEA to 96% if these cases had met eligibility for inclusion in IDEA. There is the possibility that some of these, most likely those with neurofibromatosis, tuberous sclerosis, Marfan syndrome or Prader Willi syndrome may have a milder cognitive deficit and not meet the criteria for ID.

We found one Canadian study which had used hospital morbidity codes to identify ID in at least one patient record in order to form their cohort, but had found that as many as half of the multiple records for these individuals did not code ID as a comorbidity in the hospital morbidity system.[8] It was therefore likely that other individuals with ID had been missed from their cohort due to inconsistent coding of ID as a comorbidity. The authors acknowledged that, similar to our own findings, it was likely that those who had been identified with ID were more severe. Linked data studies in New South Wales, Australia have provided further evidence of the need for multiple sources of ascertainment of ID[31] using ICD codes for ID within health datasets, as well as disability services, birth and mortality linkages to identify individuals with ID.

Practical considerations for clinical care would suggest that hospital coding which does not include reference to intellectual disability as a comorbidity may impact on the way in which service is delivered to this particularly vulnerable population. Better coding practices for ID would enable researchers to investigate directly whether care or procedures are compromised for individuals with ID and facilitate the development of ID-related policies and service planning. The hospital experiences for people with ID, who we know experience higher rates of

hospitalisation than the rest of the population,[3] have been described as relying heavily on carers for in-hospital patient assistance with failure to provide appropriate care, and lack of knowledge and discharge planning by medical staff.[32] The reliance on hospital morbidity data, as well as other administrative datasets, to identify ID in a population for research purposes has been shown to provide varied results.[9] Overall, we would not recommend that researchers use hospital morbidity datasets alone as a source of identification of intellectual disability.

Conclusion

Through linkage to a hospital morbidity dataset, this study has shown that hospital data does not adequately identify individuals with ID when compared with the population-based IDEA database. A high proportion of those uniquely identified in hospital morbidity data had died early or alternatively they had a condition not necessarily associated with ID. It is important for hospital codes to reflect the ID status of patients, primarily for the benefit of recognizing their specific needs, but also for improvement of ascertainment of ID through this source. Clearly with such a high proportion of individuals not being recognized with ID, coding practices which identify ID need to be better implemented.

Acknowledgements

The authors acknowledge the Disability Services Commission, the Telethon Kids Institute, the Western Australia Department of Education, the Catholic Education Office and the Association of Independent Schools of Western Australia for assistance with data collection for the IDEA database.

Funding Statement

This study was partly funded by an Australian National Health and Medical Research Council (NHMRC) Program Grant #572742.

Competing Interests

We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

Author Contributions

All authors contributed to the initial design of the manuscript. JB and HL were responsible for the drafting of the paper. KW was responsible for analysis and contributed to the writing of the final draft. All authors contributed to the final writing of the paper and checked for important intellectual content.

Data Sharing Statement

Data are only available through ethical approval from the Western Australian Department of Health, Human Ethics Review Committee in collaboration with the authors.

References

1. Schalock R, Borthwick-Duffy S, Bradley V, Buntinx W, Coulter D, Craig E, et al. Intellectual disability: Definition, classification, and systems of supports: American Association on Intellectual and Developmental Disabilities. 444 North Capitol Street NW Suite 846, Washington, DC 20001; 2010.

2. Oeseburg B, Dijkstra GJ, Groothoff JW, Reijneveld SA, Jansen DE. Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review. Intellect Dev Disabil. 2011;49:59-85.

3. Bebbington A, Glasson E, Bourke J, de Klerk N, Leonard H. Hospitalisation rates for children with intellectual disability or autism born in Western Australia 1983-1999: a population-based cohort study. BMJ open. 2013;3.

4. Balogh R, Brownell M, Ouellette-Kuntz H, Colantonio A. Hospitalisation rates for ambulatory care sensitive conditions for persons with and without an intellectual disability--a population perspective. J Intellect Disabil Res. 2010;54:820-32.

5. Bowley C, Kerr M. Epilepsy and intellectual disability. J Intellect Disabil Res. 2000;44 (Pt 5):529-43.

6. Fitzgerald P, Leonard H, Pikora TJ, Bourke J, Hammond G. Hospital admissions in children with Down syndrome: experience of a population-based cohort followed from birth. PloS one. 2013;8:e70401.

7. Emerson E, Hatton C. Mental health of children and adolescents with intellectual disabilities in Britain. Br J Psychiatry. 2007;191:493-9.

8. Balogh RS, Hunter D, Ouellette-Kuntz H. Hospital Utilization among Persons with an Intellectual Disability, Ontario, Canada, 1995–2001. J Appl Res Intellect. 2005;18:181-90.

9. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. Res Dev Disabil. 2011;32:419-36.

10. Büchele G, Becker C, Cameron I, Auer R, Rothenbacher D, König H, et al. Fracture risk in people with developmental disabilities: results of a large claims data analysis. Osteoporosis Int. 2017;28:369-75.

11. Petterson B, Leonard H, Bourke J, Sanders R, Chalmers R, Jacoby P, et al. IDEA (Intellectual Disability Exploring Answers): a population-based database for intellectual disability in Western Australia. Ann Hum Biol. 2005;32:237-43.

12. Holman CDAJ, Bass AJ, Rouse IL, Hobbs MST. Population-based linkage of health records in Western Australia: development of a health services research linked database. Aust NZ J Publ Heal. 1999;23:453-9.

13. Heber R. A manual on terminology and classification in mental retardation. Am J Ment Def. 1961;64.

14. Yeargin‐Allsopp M, Murphy CC, Cordrro JF, Decoufle P, Hollowell JG. Reported biomedical causes and associated medical conditions for mental retardation among 10‐year‐old children, metropolitan Atlanta, 1985 to 1987. Dev Med Child Neurol. 1997;39:142-9.

15. Bourke J, de Klerk N, Smith T, Leonard H. Population-Based Prevalence of Intellectual Disability and Autism Spectrum Disorders in Western Australia: A Comparison With Previous Estimates. Medicine. 2016;95:e3737.

16. Leonard H, Glasson E, Bebbington A, Hammond G, Croft D, Pikora T, et al. Application of Population-Based Linked Data to the Study of Intellectual Disability and Autism. In: Urbano RC, editor. International Review of Research in Developmental Disabilities. 45. MA, USA: Academic Press; 2013. p. 281-327.

17. Leonard H, Glasson E, Nassar N, Whitehouse A, Bebbington A, Bourke J, et al. Autism and intellectual disability are differentially related to sociodemographic background at birth. PloS one. 2011;6:e17875.

18. Leonard H, Nassar N, Bourke J, Blair E, Mulroy S, deKlerk N, et al. Intrauterine growth and subsequent intellectual disability in a ten-year population cohort of children in Western Australia. Am J Epidemiol. 2007;165:S19-S.

19. Bourke J, Nembhard WN, Wong K, Leonard H. Twenty-Five Year Survival of Children with Intellectual Disability in Western Australia. J Pediatr. 2017.

20. Curatolo P, Bombardieri R, Jozwiak S. Tuberous sclerosis. The Lancet. 2008;372:657-68.

21. Cassidy SB, Driscoll DJ. Prader–Willi syndrome. Eur J Hum Genet. 2009;17:3-13.

22. Hyman SL, Arthur E, North KN. Learning disabilities in children with neurofibromatosis type 1: subtypes, cognitive profile, and attention-deficit-hyperactivity disorder. Dev Med Child Neurol. 2006;48:973-7.

23. Dietz H. Marfan Syndrome Seattle (WA): University of Washington, Seattle; 2001 [updated 2 Feb, 201711 July, 2017]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1335/.

24. Baio J. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. MMWR Surveill Summ. 2014;63:1-21.

25. Nassar N, Dixon G, Bourke J, Bower C, Glasson E, de Klerk N, et al. Autism spectrum disorders in young children: effect of changes in diagnostic practices. Int J Epidemiol. 2009;38:1245-54.

26. Greer MK, Brown FR, Pai GS, Choudry SH, Klein AJ. Cognitive, adaptive, and behavioral characteristics of Williams syndrome. Am J Med Genet. 1997;74:521-5.

27. Walker CE, Mahede T, Davis G, Miller LJ, Girschik J, Brameld K, et al. The collective impact of rare diseases in Western Australia: an estimate using a population-based cohort. Genet Med. 2017;19:546-52.

28. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511:344-+.

29. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2016;17:9-18.

30. Valdez R, Grosse SD, Khoury MJ. The need for a next-generation public health response to rare diseases. Genet Med. 2016.

31. Reppermund S, Srasuebkul P, Heintze T, Reeve R, Dean K, Emerson E, et al. Cohort profile: a data linkage cohort to examine health service profiles of people with intellectual disability in New South Wales, Australia. BMJ open. 2017;7:e015627.

32. Iacono T, Bigby C, Unsworth C, Douglas J, Fitzpatrick P. A systematic review of hospital experiences of people with intellectual disability. BMC Health Serv Res. 2014;14:505.

Figure Legend:

Figure 1: Identification of intellectual disability (ID) in children born 1983-2010 and hospitalised in Western Australia using linkage to the IDEA database and the hospital morbidity data system (HMDS)

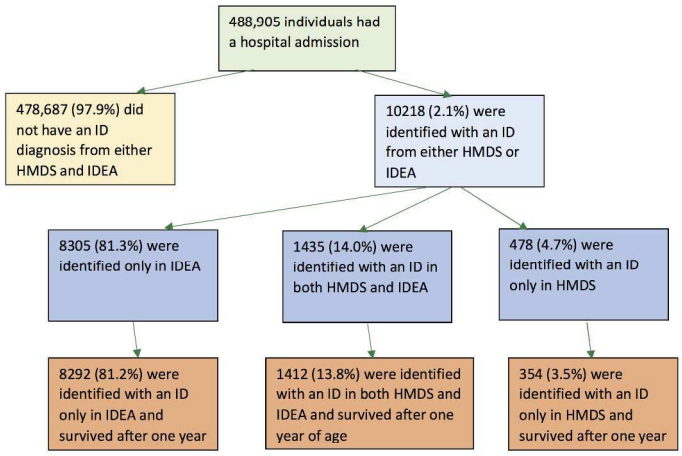

Figure 1: Identification of intellectual disability (ID) in children born 1983-2010 and hospitalised in Western Australia using linkage to the IDEA database and the hospital morbidity data system (HMDS)

209x296mm (300 x 300 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 5
Methods			
Study design	4	Present key elements of study design early in the paper	Pages 5,6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Pages 5,6
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Page 6
Bias	9	Describe any efforts to address potential sources of bias	n/a
Study size	10	Explain how the study size was arrived at	Page 5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Page 6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 6
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6
		(*c*) Explain how missing data were addressed	n/a
		(*d*) If applicable, explain how loss to follow-up was addressed	n/a
		(*e*) Describe any sensitivity analyses	n/a
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Page 7
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) Summarise follow-up time (eg, average and total amount)	n/a
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Page 7
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n/a
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 8,9
Limitations			Page 9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Pages 10,11
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.